Small RNA-directed heterochromatin formation in the context of development: what flies might learn from fission yeast

Kathryn Huisinga
Washington University in St. Louis

Sarah C.R. Elgin
Washington University in St. Louis, selgin@wustl.edu

Follow this and additional works at: https://openscholarship.wustl.edu/bio_facpubs

Part of the Biology Commons

Recommended Citation
https://openscholarship.wustl.edu/bio_facpubs/195

This Article is brought to you for free and open access by the Biology at Washington University Open Scholarship. It has been accepted for inclusion in Biology Faculty Publications & Presentations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Small RNA directed heterochromatin formation in the context of development: what flies might learn from fission yeast

Kathryn L. Huisinga and Sarah C.R. Elgin
Department of Biology, Washington University, St Louis, MO 63130, USA

Summary

A link between the RNAi system and heterochromatin formation has been established in several model organisms including \textit{S. pombe} and \textit{A. thaliana}. However, the data to support a role for small RNAs and the associated machinery in transcriptional gene silencing in animal systems is more tenuous. Using the \textit{S. pombe} system as a model, we analyze the role of small RNA pathway components and associated small RNAs in regulating transposable elements and potentially directing heterochromatin formation at these elements in \textit{Drosophila melanogaster}.

Keywords

heterochromatin; \textit{Drosophila melanogaster}; Small RNAs; \textit{S. pombe}; transposable elements; cell-cycle

Small non-coding RNA pathways

Pathways to generate small non-coding RNAs exist in a wide-range of eukaryotic organisms, from the yeast \textit{Schizosaccharomyces pombe} to plants to humans. The only known exception among the established eukaryotic model organisms is \textit{Saccharomyces cerevisiae}, which lacks the RNAi processing system, but does use antisense and non-coding RNAs to help regulate cellular processes\cite{1,2}. Small RNAs and the pathways that generate them are implicated in a variety of epigenetic processes, including control of transposable elements, genome rearrangement, RNA-directed DNA methylation, quelling, heterochromatin formation, and nucleolar dominance and are studied in a wide variety of organisms, including \textit{S. pombe}, \textit{Tetrahymena}, \textit{Drosophila melanogaster}, \textit{Caenorhabditis elegans}, \textit{Arabidopsis thaliana}, and mammalian systems. Three types of proteins are primarily involved in generating small RNAs. The first category is the RNaseIII family of nucleases, which include the Dicers and Drosha \cite{3–5}. These enzymes are responsible for cutting longer double-stranded RNA (dsRNA) into smaller fragments of 20–30nt. Second are the dsRNA binding (DRB) proteins, which act as partners to the RNAseIII enzymes, and include loquacious (loqs), R2D2, and Pasha in \textit{Drosophila}\cite{6–11}. The third category is the Argonaute family of PAZ/Piwi domain containing proteins, which bind to the small dsRNAs and utilize that information to direct a variety of gene silencing effects. The PAZ domain is important for binding the small RNA, while the Piwi domain mediates the endonucleolytic cleavage, aka “slicer” activity, responsible for cutting the target RNAs identified by the small bound RNA. The Argonaute family proteins

\footnote{1Corresponding author e-mail: khuising@wustl.edu, fax: (314) 935-4432.}

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
are divided into two clades, the Piwi clade and the Ago clade, each having distinct functions as discussed below. Metazoans contain members from both clades, while the sole Argonaute in fission yeast and those in plants are all from the Ago clade [reviewed in 12].

To date, there have been three small RNA pathways described in animals (see Figure 1) [reviewed in 13]. The siRNA pathway is most familiar, as this is the pathway employed in RNAi knockdown experiments. This pathway processes long dsRNA using a Dicer enzyme (Dicer-2 in Drosophila) to generate short 21–23nt dsRNAs that are incorporated into an Argonaute family member-containing siRNA-induced silencing complex (siRISC, containing AGO2 in Drosophila). This complex utilizes the slicer activity of its Argonaute to destroy mRNA transcripts having sequence complimentary to the small RNAs. In flies, as in plants and worms, the endogenous function of this pathway includes acting as a viral defense system, as mutations in this pathway result in increased susceptibility to viral infection [reviewed in 14]. Very recent publications have implicated components of this pathway (Ago2 and Dcr2) in the generation of endogenous siRNAs (esiRNAs) derived from transposable elements, convergent transcripts, and long hairpin RNA genes in Drosophila embryos and somatic cells [15–20]. These papers are discussed in a recent review [21]. The details of this pathway and how it interfaces with the other pathways in Drosophila remain to be elucidated. The miRNA pathway uses RNA polymerase II (RNAPII) transcribed RNAs that fold back on themselves to form imperfect hairpins that are the source of dsRNA. In flies, these hairpins are processed first by the Drosha/Pasha complex and then by the Dicer-1 endonuclease and its DRB partner, losq, to create short 21–23nt RNAs. These RNAs are loaded into the AGO1 containing miRISC, which blocks translation of mRNAs with imperfect pairing to the miRNA present in the complex. Depending upon the amount of base-pairing between the miRNA and the target mRNA, slicing of the mRNA can occur as well. The primary function of this pathway appears to be fine-tuning the expression of genes in a developmental context [reviewed in 22]. The third small RNA pathway present in animals is the piRNA pathway in which members of the Piwi clade of Argonaute proteins produce somewhat longer small RNAs (24–30nt) generated (at least in part) from specific genomic regions, termed piRNA clusters. Repeat associated siRNAs (rasiRNAs) identified in Drosophila are one class generated by the piRNA pathway [23]. piRNA clusters in flies are composed primarily of imperfect copies of transposable elements, including both retrotransposons and DNA transposons [24]. In mice there are different piRNA clusters expressed at different developmental stages, some of which are enriched for transposable elements, some of which are not (see below). The piRNA pathway appears to be critical for maintenance of germline stability, as mutations in this pathway result in fertility defects [25–29].

In this review we focus on the proposed role of small RNAs and of the proteins they are associated with in directing and maintaining the formation of heterochromatin. This relationship is well established in S. pombe [30], which serves as our paradigm for discussing the possible mechanisms. Often this function of small RNAs and the RNAi machinery is referred to as transcriptional gene silencing, as the end result is to shut off transcription via modification of chromatin. However, in the best-understood system, S. pombe, there is a post-transcriptional aspect to this process in that the regions of the genome destined to be transcriptionally silenced (i.e. the dg-dh pericentromeric repeats and the MAT locus) are actually transcribed during S-phase of the cell-cycle (see below). Here, we will consider the mechanisms that ultimately result in changes in chromatin structure with subsequent decreases in gene expression at a transcriptional level, as transcriptional gene silencing, while acknowledging that transcription of the silenced region and subsequent processing of these transcripts, can be involved in this process. One can, however, differentiate this mechanism from straightforward post-transcriptional gene silencing mechanisms, which function solely by degrading transcripts and do not result in any changes in chromatin structure or in the transcription level of the homologous region. In addition to S. pombe, links between small
RNAs and heterochromatin formation have been detected in several animal models, including flies, zebrafish, and mammalian systems. We review what is known about the relationship between small RNAs and heterochromatin formation and examine some unanswered questions about this relationship by focusing on the S. pombe and Drosophila systems, but include a brief examination of the links in the mammalian systems. Other recent reviews have examined the roles for small RNAs in epigenetic regulation in other model organisms in more detail [12, 31–33].

**Formation of heterochromatin in S. pombe utilizes cell-cycle regulated small RNAs**

Heterochromatic regions of the genome in the model organism S. pombe include the silent mating type (MAT) loci, subtelomeres, the ribosomal DNA (rDNA), and the regions surrounding the centromeres (pericentric domains)[34]. Reporter genes located in these regions exhibit a variegating phenotype, characteristic of heterochromatin[35]. The chromatin of these regions is enriched for Swi6 [the S. pombe Heterochromatin Protein 1 (HP1) homologue], histone 3 methylated at lysine 9 (H3K9me), and Clr4, the S. pombe homologue of the Su(var) 3–9 histone methyltransferase [34]. All of these biochemical marks are associated with heterochromatin. A long-standing question is how these marks are targeted to the specific regions of a genome, which are appropriately heterochromatic, but not to other regions. Much work in fission yeast over the past several years has elucidated a role for the RNAi machinery and small RNAs in directing the formation of heterochromatin in the pericentric regions, where this pathway is required for heterochromatin formation[36]. The RNAi pathway is also required for establishment of heterochromatin at the silent mating-type loci, but here it plays a redundant role with the ATF/CREB pathway in maintaining the heterochromatic state[37,38].

Small RNAs (20–22nt) with identity to an endogenous heterochromatic sequence, the dh repeats at the pericentromere, were the first identified in S. pombe[39]. Subsequent characterization of the S. pombe small RNA population has identified not only RNAs with sequence homology to the repetitive dg and dh elements present at the outer (otr) regions of S. pombe centromeres, but also RNAs matching the inner (inr) centromeric repeats, the CenH region of the MAT locus, the rDNA and the subtelomeres[34,39]. A self-reinforcing feedback loop model for the generation of these small RNAs, suggesting how the RNAi processing machinery works to recruit heterochromatin factors to the pericentromere, has been proposed (Figure2)[40,41]. Transcription of the dg-dh repeats in the otr region of the centromeres is dependent on RNAPII, as mutations in subunits of RNAPII result in a loss of small RNAs and a loss of silencing[42,43]. At the dg-dh repeats, both the forward and reverse strands can be transcribed[36]. The reverse strand is transcribed even in the presence of heterochromatin, but transcription of the forward strand is repressed transcriptionally by Swi6 recruitment[36,42]. Ago1, the sole Argonaute family protein in S. pombe, interacts with the polymerase and plays a role in processing these transcripts via its slicing activity[44,45]. Mutations which disrupt the catalytic activity of Ago1 result in disruption of heterochromatin formation and the loss of spreading of heterochromatin into reporter genes inserted in the otr region[46]. Ago1 is a component of the RNA-induced transcriptional silencing (RTS) complex, which also includes Chp1, a chromodomain-containing protein which is dependent upon Clr4 mediated H3K9 methylation for localization to heterochromatic regions[47], and Tas3, a protein that appears to play a bridging role between Ago1 and Chp1 in the complex[48–50]. In addition, the RTS complex contains small RNAs with homology to the dg-dh repeats[50]. RTS interacts with a second complex, the RNA-directed RNA polymerase complex (RDRC), which contains RdP1, an RNA-dependent RNA polymerase (RdRp) that uses the RNAPII transcript as a template to generate dsRNA[40,51]. RDRC contains two additional components, Cid12, a member of the polyA polymerase family of enzymes, and Hrr1, a putative helicase[51]. The Dicer enzyme interacts with RDRC[52] and processes the dsRNAs produced, thereby generating additional
small RNAs, which can then feed back into the RITS complex to be bound by Ago1. A second Ago1 containing complex, Argonaute siRNA chaperone (ARC), plays a role in this process by regulating the conversion of the small double-stranded RNA generated by Dicer into the single-stranded guide RNAs present in RITS[53]. The two additional proteins in ARC, Arb1 and Arb2, appear to play a role in regulating this conversion[53]. This process is thought to occur in cis at the pericentromeric region being transcribed. Chromatin immunoprecipitation experiments have demonstrated that many of these factors (RNAPII, RITS, and RDRC) are present at the silenced heterochromatic regions in S. pombe[34,41,43,47]. Evidence for a cis mechanism comes from the observation that tethering the RITS complex (via a Tas3-AN peptide fusion protein which binds to its cognate RNA binding site, the BoxB) to the nascent transcript of a ura4 reporter gene drives formation of heterochromatin and transcriptional gene silencing, but only in cis, not in trans[54]. This cis restriction is regulated by the ribonuclease Er1 (which presumably degrades excess siRNA), as Er1 deletion strains are able to silence a ura4 reporter in trans as well as the ura4 reporter where the Tas3-AN peptide is tethered[54].

Small RNA processing is required for the formation of heterochromatin at the centromeric repeats in S. pombe; deletion of components of the RNAi pathway results in the loss of silencing of reporters inserted into this region and the loss of heterochromatic marks (i.e. Swi6 & Clr4 binding, H3K9me)[36,50]. It is proposed that the targeting of RITS, via the small RNAs with homology to the repeats, promotes the binding of Chp1 to methylated H3K9[50] (see Figure 2). However, the recruitment of the RNAi machinery and the Clr4 histone methyltransferase (HMT) which deposits this mark at the pericentromere are co-dependent, in that loss of one results in the loss of the other[34,40,41,55]. Additional insight into Clr4 recruitment comes from the observation that Clr4 exists in a protein complex, designated CLRC, with several additional factors including Rik1, a WD repeat containing protein that has been proposed to help recruit Clr4 to heterochromatic regions[56–58]. Also present in the complex are the cullin protein Cu4, a component of the cullin-dependent ligase (CDL) family of E3 ubiquitin ligases, and two addition components of the CDL family, Ned8, a neddylation factor which modifies cullins, and Pip1, a Ring finger protein[56,57]. Two additional proteins, Dos1/Clr8 (Raf1/Cmc1) and Dos2/Clr7 (Raf2/Cmc2), which were identified in screens for mutants that disrupt Swi6 localization and heterochromatin formation, are also in the CLRC complex[56,57,59,60]. Loss of components of the CLRC complex results in disruption of heterochromatin formation at the pericentromere. A recent study has shown that recruitment of the Rik1 component of CLRC is dependent upon the RNAi pathway[55]. Mutations that disrupt Ago1 or Rdpl activity result in the loss of Rik1 recruitment. Mutations in Cu4 disrupt the recruitment of Clr4, but not of Rik1, indicating a potential role for Cu4 in bridging a Rik1-Clr4 interaction. The binding of Rik1 correlates with increased transcription of repeat regions and coincides with RNAPII occupancy, lending support to the model that Rik1 plays an upstream, targeting role in the recruitment of Clr4. Therefore, it appears that the Rik1 subunit of CLRC mediates recruitment of the complex in an RNAi dependent manner. In addition, CLRC interacts with the RITS complex in a manner dependent upon the Tas3 subunit[55], which may also facilitate recruitment of CLRC to heterochromatic regions. Taken together, this data supports a model where the recruitment of CLRC and RITS are co-dependent.

A puzzling aspect of this model is exactly how RNAPII dependent transcription of the pericentric repeats occurs in a heterochromatic region, as heterochromatin packaging is generally considered inhibitory to the RNAPII transcription machinery[reviewed in 61,62]. Recent publications showing that RNAi-mediated heterochromatin formation appears to be regulated in a cell-cycle dependent manner, with transcription of the heterochromatic regions occurring during S-phase when the DNA is not yet completely packaged as heterochromatin, help to explain some of the apparent contradictions[63,64]. As cells enter mitosis, H3K9me and Swi6 binding are reduced; this reduction persists until S phase. H3K9me and Swi6 binding then gradually increase again until G2 when binding is maximal (Figure 2). There is an inverse
relationship between Swi6 binding and the presence of histone 3 phosphorylated at serine 10 (H3S10ph), which is highest during mitosis. The increase in H3S10ph coincides with the recruitment of condensin, which is implicated in mitotic chromosome condensation. Binding of RNAPII to the heterochromatic regions and the generation of heterochromatic transcripts occurs during S phase[63,64]. However, in a clr4Δ strain, the binding of RNAPII increases throughout the cell cycle, indicating that heterochromatin assembly does normally limit RNAPII occupancy as one would predict[63]. Rik1, which is critical for loading CLRC, also shows peak binding during S phase, coinciding with the RNAPII peak, whereas RITS binding increases during S phase and persists into G2, consistent with the data showing that Ago1 interacts with RNAPII. Clr4 and the CLRC component Raf2 show profiles similar to H3K9me and Swi6, with the least binding during mitosis, gradually increasing during S phase, with the peak of binding during G2. Concurrent with transcription by RNAPII during S phase is an enrichment of H3K36me and the Alp13/Eaf3 protein[63]. Analysis of RNAPII transcription at euchromatic genes has demonstrated that RNAPII targets the H3K36 methyltransferase Set2 to the transcribed region[65], which in turn recruits histone deacetylase complexes (HDACs) via interaction of the Alp13/Eaf3 chromodomain with the H3K36 methylated histone[66,67]. This recruitment presumably facilitates reassembly of chromatin following transcription. The report of H3K36me and Alp13/Eaf3 at the heterochromatic repeats indicates that a similar process is used when RNAPII is transcribing heterochromatic regions. These findings indicate that S phase is permissive for transcription of heterochromatic regions, but that the transcripts are processed to small RNAs by recruitment of Ago1 and other components, which then facilitate the recruitment of the heterochromatin machinery that is present at these regions during G2. These finding have significant implications for understanding heterochromatin formation and stability in multicellular organisms, which contain different cell types that are undergoing division at different rates or are terminally differentiated and have exited the cell cycle.

A key question in understanding the function of heterochromatin formation in the centromere region is the relationship between pericentric heterochromatin and the functional centromere. Recent work using a minichromosome transformation assay has added to the understanding of this relationship by demonstrating that heterochromatin proteins Swi6, Chp1, and Clr4, and the RNAi machinery, in particular Dicer, are necessary to promote CENP-A deposition and kinetochore assembly over the central domain of a fission yeast centromere on a naïve template[68]. This appears to be an absolute requirement only for establishing the centromere, as loss of these factors does not affect CENP-A deposition on endogenous centromeres or on minichromosome centromeres which have already incorporated CENP-A in a previous generation. Consistent with the requirement for heterochromatin, the otr region (which contains the dg-dh repeats), in addition to the cnt region where CENP-A binds, must be present on the naked template[68]. These results are in agreement with the previous observation that cells deleted for Ago1, Dicer or Rdp1, while viable, have defects in chromosome segregation[69, 70]. Interestingly, mutations in some of the Drosophila RNAi components with a potential role in heterochromatin formation exhibit defects in chromosome segregation as well as in gene silencing (see below).

**Generation of rasiRNAs in Drosophila melanogaster**

Cloning and sequencing Drosophila small RNA populations over a developmental time course first identified the rasiRNA class of small RNAs in flies. These small RNAs, which can be considered a subclass of the larger piRNA class of small RNAs, are ~24–30nt long and have homology to repetitious sequences including transposable elements[71]. Additional characterization of the rasiRNAs has shown that they have only a 5′ phosphate group; based on their resistance to periodate oxidation/β-elimination, it was concluded that they lack either a 2′ or 3′ hydroxyl group at the 3′ terminus[72]. Subsequent studies have shown that the 3′ end
of rasiRNAs is 2′-O-methylated by the Drosophila homologue of Arabidopsis thaliana HEN1, DmHen1/Pimet[73,74]. This class of small RNAs is primarily present in the ovaries, testes, and early embryos of flies, although a few RNAs from this class can be isolated from later stages of development[71]. Analysis of two specific types of rasiRNAs, those homologous to the Suppressors of Stellate locus (which controls expression of the X-linked Stellate gene in an RNAi dependent manner[75,76]) and those homologous to the LTR-containing retrotransposon roo (found in the germline of adult flies), demonstrated that the generation of these rasiRNAs in ovaries or testes is independent of several key genes in the siRNA pathway, including Dcr-2, its DRB partner R2D2, and Ago2, as well as loqs, encoding the DRB protein required to produce miRNAs[72]. Dependency upon Dcr-1 was more difficult to examine, as flies homozygous null for Dcr-1 are not viable. Nevertheless, using mitotic mutant clones in ovaries, it was demonstrated that Dcr-1 as well is not required for production of the roo rasiRNA. The lack of dependency upon Dcr-1 and Dcr-2 is consistent with the fact that these enzymes previously had been shown to produce small RNAs of ~22nt[3]. Subsequent studies showed that the presence of wild type levels of these rasiRNAs is dependent upon two different members of the Piwi clade in the Argonaute family, piwi and aubergine (aub), as well as two putative RNA helicases, spn-E (aka hls), a member of the DE-H family of RNA-dependent ATPases[77], and armitage (armi), a homologue of the Arabidopsis SDE3 protein that contains an ATP-dependent domain distinct from DEA(H/D) box helicases[72,78]. Previous studies have demonstrated a role for armi in assembly of the siRISC complex from ovaries[79] and a requirement for spn-E in Stellate silencing[80] and RNAi activation during oocyte maturation[81].

Further understanding of how rasiRNAs are generated and what their biological role might be has come from sequencing the small RNAs associated with the Argonaute proteins from the Piwi clade in flies, Piwi, Aub, and AGO3. Purification and subsequent sequencing of the small RNAs associated with these proteins in ovaries or testes has shown that the large majority share identity with transposable elements (including both class I retrotransposons and class II DNA transposons) and other repetitious sequences present at heterochromatic loci[24,82–85]. Using the rasiRNAs that map to only one genomic location (i.e. uniquely mapping rasiRNAs), several hundred discrete loci where rasiRNAs are produced have been identified[24,85]. These loci are termed “piRNA clusters” or “rasiRNA clusters” and vary in size from several Kb to hundreds of Kb. The majority of these clusters are located in the β-heterochromatin present at pericentric regions, the fourth chromosome, and at telomeres, where there are numerous partial, inactive transposable elements. These clusters account for more than 92% of all of the sequenced piRNAs[24]. The largest cluster, located at cytological region 42AB, has homology to ~30% of all of the sequenced piRNAs and is 240kb in length[24]. Other prominent clusters include telomeric clusters on most arms (X, 2R, 2L, 3R and 4) that have homology to both telomere associated sequence (TAS) repeats and the Het-A and TART transposons, and a cluster present at 20A (flamenco, discussed in more detail below) which previously had been identified as a region involved in transposable element (TE) control[86,87]. It appears that there is a high level of complexity in piRNAs, as the three sequencing projects which examined Piwi-associated RNAs from ovaries show little overlap among the actual sequences of the small RNA populations[85], although these studies all identified clusters in many of the same genomic regions.

When the sequences of the rasiRNAs associated with the three different Piwi clade Argonaute proteins are compared, a strong strand-specific asymmetry is observed. Piwi and Aub are associated with antisense strands while Ago3 is found with sense strands, although this asymmetry appears to be reversed for some families of TEs. Upon examining the relationship between these two classes of rasiRNAs, one observes that Aub and Piwi associated rasiRNAs exhibit a strong bias for U at the 5′ end, while AGO3 associated rasiRNAs have a preference for A at nucleotide 10. Based upon this sequence bias and comparisons between the cloned
sequences of the different classes, a “ping-pong” model for the generation of these small RNAs has been proposed (Figure 3) [24,82]. In this model, a primary piRNA is generated by transcription of the piRNA cluster. This primary piRNA transcript is then processed by a mechanism that uses rasiRNAs from the transposable elements oriented in the antisense direction which are bound by Piwi or Aub. These Piwi/Aub/RNA complexes pair with sense transcripts, possibly generated from active transposons or potentially from TEs oriented in the sense direction in piRNA clusters, and guide the slicing of the 5′ end of the sense transcript. (The mechanism for generating the 3′ end is currently unknown.) The sense piRNAs are bound by AGO3 and can then pair back with antisense transcripts from the piRNA clusters and other sources and guide the slicing of their 5′ end. These antisense piRNAs are bound again by Piwi/Aub and can be used to continue the cycle. In addition to the strand bias, the cellular localization of these three Argonaute proteins in ovaries also differs. While Piwi is localized in the nucleus of both nurse cells and follicle cells, Aub and AGO3 are located in the cytoplasm, and show enrichment in the nuage, an electron-dense perinuclear structure present in the germline of many animals. In flies, many factors involved in small RNA processing are localized in the nuage (Table 1). All three proteins are loaded into the developing oocyte, as shown by the fact that all are present in 0–2h embryos; Aub shows an accumulation in the posterior pole, where the primordial germline forms, of stage 10 developing oocytes and persists there, being found exclusively in germ line cells in the embryo[24]. Interestingly, the greatest amount of complementarity between rasiRNAs is seen between those bound to AGO3 and Aub, both of which localize to the nuage.

While the ping-pong model explains much of the data, in particular how small RNAs are generated independent of Dicer proteins, many unanswered questions remain. It is unclear how the 3’ ends of rasiRNAs are generated. The model requires that the sliced rasiRNA is cleaved by a yet-to-be-identified nuclease. Potential candidates for this nuclease include the products of *zucchini* (*zuc*) and *squash* (*squ*), which have been identified to play a role in rasiRNA production and localize to the nuage[88]. However, whether or not these or other nucleases are used remains to be determined. A more complicated question is how the strand bias is achieved among the different Argonaute proteins. Most piRNA clusters are bidirectionally transcribed, generating a mix of sense and antisense transcripts[23,24]. A mechanism by which an active TE, which would presumably be transcribed in the sense direction, could be biased for regulation by one member of the Piwi clade can be envisioned by potential co-recruitment of an Argonaute family member with the transcription machinery. As mentioned previously, in fission yeast Ago1 interacts with RNAPII [44,45], and studies of the plant RNA Pol IV polymerase found interactions between RNAP IV and Argonaute 4 through the functionally conserved WG/GW motif[89,90]. A more complicated mechanism would need to be invoked for establishing the strand bias for transcripts generated from piRNA clusters, where the model proposes the generation of one long transcript encompassing the entire cluster[24]. In the case of the *flamenco*/*COM* locus at cytological band 20A, most of the TEs within the cluster are oriented in the antisense direction, but this bias does not occur in the majority of piRNA clusters, where partial and degenerate TEs are oriented in both the sense and antisense direction [23,24]. In this circumstance, the co-recruitment of an Argonaute family member with the polymerase, while quite possible, is insufficient to completely explain the strand bias. Finally, the source of the sense transcripts used to initiate the process has not been fully elucidated. Full-length active TEs, either located in euchromatin or buried in heterochromatin, or partial TEs, located in piRNA clusters and oriented in the sense direction, could potentially give rise to such RNAs.
Role of rasiRNA machinery in silencing transposable elements in the germline

Investigations into rasiRNAs and the proteins involved in generating rasiRNAs have identified a role for these factors in regulating the expression and transposition of transposable elements in the male and female germlines. In terms of this discussion, it is helpful to clarify the definition of “germline”. In the strict sense, this term refers solely to the cell lineage producing the gametes. However, experimentalists in Drosophila have used the term to include all types of cells present in ovaries or testes. Thus in the case of the ovary, this definition encompasses the oocyte, the nurse cells, and the surrounding follicle cells; the latter are actually a somatic cell type, although they play a critical role in germline function. In-depth analysis of the piRNA cluster located at cytological band 20A has demonstrated that mutations disrupting the locus cause up-regulation and mobilization of the TEs sharing sequence identity with the locus. Previous studies isolated mutations in region 20A as flamenco(\textit{flam})/\textit{COM}, a controlling region for mobilization of the LTR-containing retrotransposons gypsy, \textit{Zam}, and \textit{Idefix}. gypsy is a retrovirus that is expressed in follicle epithelium, where it is packaged into a virus-like particle and can invade the ovary, while \textit{Zam} is expressed in posterior follicular cells, and \textit{Idefix} is expressed in the gerarium.

Regulation of mobilization of these transposons is defined by two different genetic states; one allows mobilization [referred to as Permissive (P), in the case of gypsy regulation, or Unstable (U) in the case of \textit{Zam} or \textit{Idefix} regulation], and one that does not allow mobilization [classified as Restrictive (R, for \textit{gypsy}) or Stable (S, for \textit{Zam}/\textit{Idefix})]. Mutations have been identified which cause a switch from an R strain to a P strain, leading to increased accumulation of gypsy RNA in the follicle cells, loss of control of gypsy and \textit{Zam} mobilization (\textit{Idefix} was not tested in this assay), and subsequent female sterility[87,91]. These mutations were mapped to region 20A; upon identification of this locus as a piRNA cluster, it was demonstrated that these \textit{flam} mutants had reduced levels of piRNAs that map to the 20A cluster and that the generation of the primary piRNA transcript had been disrupted[24,91,92]. Additional studies demonstrated that mutations in \textit{piwi}, the only Argonaute family member present in the follicle cells, also caused a global decrease in gypsy rasiRNAs and a corresponding increase in expression of gypsy transcripts in the germline[92,93]. A second body of work, focused on the regulation of \textit{Zam} and \textit{Idefix}, demonstrated similar effects of up regulation and mobilization of these TEs in response to mutations in \textit{flam}/\textit{COM}, and a dependency upon \textit{piwi} for control of these elements in the germline[86,94]. Together, these studies have linked three observed effects: change in the level of rasiRNAs with homology to a TE, increased expression of the TE in the germline, and mobilization of the TE in the germline. Either mutations in \textit{flam}/\textit{COM}, which disrupt the primary piRNA transcripts, or mutations in \textit{piwi}, which appears to be required for processing these transcripts, can cause these three effects. The results suggest that the piRNA system limits the presence of TE transcripts, which in turn will limit transposition. Clearly there is a component of this mechanism which utilizes post-transcriptional processing of the TE transcripts. It remains to be determined whether this part of the shut-down mechanism then directs the changes in chromatin structure which occur at these regions, similar to what is observed in \textit{S. pombe}, or whether another separate mechanism directs the changes in chromatin structure, while the ping-pong mechanism is only concerned with removing TE transcripts in the germline by post-transcriptional gene silencing (i.e. RNA degradation). If this post-transcriptional silencing of the TEs is linked to the chromatin modifications of these regions, this would imply that the \textit{cis} requirement observed for processing transcripts in \textit{S. pombe} does not apply in Drosophila, given that some of the ping-pong pathway components involved in processing the rasiRNAs are not even present in the nucleus (i.e. Aub & AGO3). Additional evidence that the \textit{cis} requirement may not apply in flies can be found in the fact that loss of \textit{Snipper}, the putative Drosophila homologue of the \textit{Eri1} exonuclease (which imparts
Additional studies have demonstrated a role for other components of the RNAi machinery in regulating the expression of a variety of different TEs and in the generation of rasiRNAs in the germline. Table 1 summarizes the effects observed in the ovaries on the expression of a subset of TEs in the presence of mutations in different factors linked to regulation of the rasiRNA pathway. These include many germline-specific factors that localize to the nuage, such as the products of *aub*, *krimper*, *maelstrom*, and *cutoff*, as well as the putative nucleases, *zuc* and *sqa*, mentioned previously[72,88,96–98]. Also included are *spn-E* and *armi*, two helicase family members that have been implicated in a variety of roles in RNAi-based regulation[79, 81]. As there are no mutations available in the AGO3 gene, a requirement for AGO3 in regulation of TE expression has not been tested. The effect of mutations in the different components varies to some extent depending upon which TE is examined. This variability may indicate a division of labor within the system and/or may reflect cell specific differences. However, no up-regulation of TEs in the female germline is observed with mutations in the siRNA pathway components AGO2, Dcr-2, or R2D2, or in the miRNA pathway component *loqs*, in agreement with their lack of effect on rasiRNA production[72].

In addition to increased TE expression, many lines with mutations in rasiRNA components are female sterile and show defects in dorsal-ventral patterning of the embryo. Investigation using *armi* and *aub* mutants has found that these embryonic patterning defects can be suppressed by mutations in *mei-41* and/or *mnk*, which encode the fly ATR and Chk2 kinases activated by DNA double-strand breaks (DSBs)[99]. In agreement with this result is the observation that γH2A.v foci are present in the germline of the *armi* and *aub* mutants. (γH2A.v is a phosphorylated form of the H2A histone variant that is associated with DNA DSBs.) While the ATR/Chk2 mutants suppress the patterning defects, they do not alter the defects in TE silencing associated with mutations in these rasiRNA pathway members. A similar suppression of patterning defects is observed between *mnk* mutations and *cuff* and *sqa* mutations; however, there is no suppression of patterning defects by *mnk* mutations in *zuc*, *spn-E* or *piwi* mutants[88,97,99]. It has been proposed that this suppression of patterning defects occurs due to the bypass of the DSB checkpoint that would normally be activated when TE expression is up-regulated and uncontrolled mobilization of TEs occurs, resulting in DSBs throughout the genome. The lack of suppression with *zuc*, *spn-E*, and *piwi* mutations indicates an additional role for these factors in a pathway outside of that mediated by Chk2.

### Relationship between transposable element silencing and heterochromatin formation in the germline

There is significant overlap in the Drosophila genome between the location of TEs and the packaging of DNA into heterochromatin, as defined by the presence of HP1 and H3K9me2, and as assessed by the recovery of variegating reporter inserts. While there is abundant evidence that rasiRNAs and the associated processing machinery regulate the expression of transposable elements in the germline, this process might be based solely on a post-transcriptional gene silencing mechanism that is unrelated and independent from the chromatin modifications in these regions. However, control of TE expression and mobilization could also occur via a transcriptional gene silencing mechanism, potentially similar to that used in *S. pombe*, where transcription of a repeat region helps to drive the formation of heterochromatin. Based on the *S. pombe* model and the data available from flies, one can propose a model in which the rasiRNAs function to direct the heterochromatin machinery (HP1, HMTs, and other factors) to these regions in a manner that is dependent upon transcription through the rasiRNA cluster to establish heterochromatin in the germline. As discussed above, data from *S. pombe* indicate...
that this transcription of the heterochromatin regions occurs predominantly during S phase of the cell cycle[64]. There are interesting implications for the role of transcriptional gene silencing in the Drosophila germline, since while the oocyte is going through meiosis germline nurse cells undergo endoreduplication. In addition to the piRNA clusters, partial (presumably inactive) TEs are scattered throughout heterochromatic domains in Drosophila, including regions in pericentric heterochromatin and on the fourth chromosome, which have not been implicated as piRNA loci. It is unclear whether the piRNA system can impact heterochromatin formation at these partial TEs and if so, whether there are transcripts through these regions, or if this impact occurs through a separate mechanism. The possibilities are intriguing.

The strongest evidence for a transcriptional gene silencing mechanism in the germline comes from studies with mutations in \( spn-E \) in the ovaries of adult flies. Chromatin immunoprecipitation assays have demonstrated that mutations in \( spn-E \) affect chromatin structure at three different endogenous transposable elements (\( HeT-A \), \( I \) element, and \( copia \)) in ovarian tissue isolated from homozygous females. A decrease in the repressive marks H3K9me2, H3K9me3 and HP1 and an increase in the active mark H3K4me2 are observed at both the promoter and ORF regions of these elements[100]. These observations are the first to actually link a mutation in a component of the rasiRNA pathway, with its associated decrease in rasiRNAs, to a change in chromatin structure and an increase in TE expression, all in one (complex) tissue type, the ovary.

In spite of this observation, other data indicate that this mechanism may be more complicated than a simple extrapolation from the case of \( S. pombe \) would suggest. The chromatin structure changes reported above are not seen in adult somatic cells of \( spn-E \) mutations (as assayed in the “carcass” tissue, the remainder of the fly after dissecting out the ovaries), suggesting that once set, heterochromatin can be maintained in the absence of RNAi in differentiated somatic tissues. Further, mutations in \( Su(var)2–5 \) (HP1a) or \( Su(var)3–9 \) (H3-K9 HMT), two key components of heterochromatin formation in flies, have no effect on the expression of the \( copia \) or \( mdg1 \) TEs in testes[101]. While this data could indicate a post-transcriptional gene silencing mechanism, differences in the chromatin system in the germline cells may also be critical. Possible explanations for this result are the presence of a testes specific HP1 variant, HP1e[102], which may be functionally redundant to HP1a. There is also an ovary specific HP1, HP1d/Rhino, whose possible role in heterochromatin formation in the ovary is largely unexplored[102,103]. Furthermore, there are at least three active HMTs in Drosophila[104]; while \( Su(var)3–9 \) was the first identified, it does not perform all of the critical functions of histone H3 K9 methylation as shown by the fact that flies with null mutations in \( Su(var)3–9 \) are viable and fertile. This survival is distinct from mutations in \( Su(var)2–5 \), which are larval lethal[105], and mutations in rasiRNA pathway components, which are generally female sterile and/or maternal effect lethal (see Table 1 and references therein). While \( Su(var)3–9 \) mutations impact heterochromatin formation, as assayed by position effect variegation, in the pericentric regions, they do not cause loss of silencing at the primarily heterochromatic 4\(^{th}\) chromosome[106]. In contrast, the \( dSETDB1/Eggless \) HMT, which plays an important role in oogenesis, also has a major role in regulating heterochromatin formation on the 4\(^{th}\) chromosome, as well as in other regions of the genome[107–110]. As high levels of \( dSETDB \) are present in testes [109], it is possible that this HMT plays the primary role in H3K9 methylation and heterochromatin formation in the male germline, which would explain the lack of effect on TE expression in \( Su(var)3–9 \) mutants. However, a functional role for either \( dSETDB \) or the third Drosophila HMT, G9a, in the male germline has not yet been explored.

Studies on the mobilization of \( HeT-A \) and \( TART \) elements, which are two non-LTR retrotransposon-like elements arrayed at Drosophila telomeres, may also help to elucidate the relationship between components of the rasiRNA pathway and components of the heterochromatin pathway. Flies heterozygous for mutations in the \( spn-E \) or \( aub \) genes have an
increased frequency of transposition of HeT-A and TART elements to broken chromosome ends [111]. Upon further analysis of the newly repaired chromosomes, it was discovered that the majority of the attachments were a TART element and that some of the repairs from the same cross have identical breakpoint junctions, indicating that the repair may have taken place at a premeiotic stage of oogenesis. This increase in transposition corresponds with the greater increase in TART expression relative to HeT-A in the ovariies of heterozygous spn-E mutants, although both TEs show strong up-regulation in the ovariies of homozygous aub or spn-E mutant flies[111]. The same assay had been used previously to demonstrate a role for HP1α in regulating the elongation of telomeres[112]. Mutations in Su(var)2–5 result in an increased frequency of HeT-A or TART attachment to the broken chromosome end, again corresponding to an increased level of HeT-A transcript in RNA isolated from female Su(var)2–5 mutant flies indicating a role for HP1α in regulating expression and transposition of HeT-A and TART in the female germline[112]. However, these studies were not able to distinguish whether the repair of the broken end occurs through an increase in HeT-A or TART transposition to that end, or by recombination with other telomeres.

A role for the rasiRNA pathway in heterochromatin formation in somatic cells?

The links between heterochromatin and “RNAi” become less clear when one examines whether or not the rasiRNA pathway plays a role in heterochromatin formation outside of the germline. Examination of the literature covering the role of rasiRNA pathway members in somatic tissues leaves one with some apparently contradictory observations. Unfortunately, maternal loading of both proteins and RNA into the oocyte blurs the line between germline and somatic cells, making genetic dissection between these two cell types difficult in Drosophila. In this section, we attempt to address some of these apparent contradictions and examine the role of rasiRNA pathway members in the context of the development of the organism.

In examining how rasiRNA pathway components may be involved in heterochromatin formation, one must consider the developmental regulation of the process. Establishment of heterochromatin in the somatic tissues of the fly apparently begins in the early embryo during syncitial blastoderm. HP1 is enriched in the intensely DAPI-staining regions corresponding to heterochromatic regions of the genome, as first visualized at ca. nuclear replication cycle 10, in nuclei located at the apical surface of the embryo[113]. This enrichment increases significantly during nuclear cycle 14 when cellularization of the blastoderm occurs[113]. Heterochromatic silencing of a lacZ reporter gene is initiated at the onset of gastrulation, around stage 6, which occurs within nuclear cycle 14[114]. A high degree of silencing is maintained through the mitotic cycles of larval development, until a relaxation occurs at the end of the 3rd instar larval stage (associated with the onset of differentiation of adult tissues), leading to a variegating phenotype[114]. The pattern of silencing observed after this relaxation during 3rd instar correlates with the pattern of silencing in the adult organism[115]. The presence of heterochromatin in the organism must be considered in the context of these stages, in that heterochromatin formation must be initiated correctly in the early embryo and maintained in order to observe a heterochromatic-responsive read-out in observations of the late larval or adult fly tissues. Mutations that disrupt the establishment, as well as mutations which fail to maintain heterochromatin or which cause misregulation of the relaxation stage, would be expected to affect the read-out observed. Given that robust zygotic transcription occurs starting in the interphase of nuclear cycle 14[116] (albeit some zygotic transcripts are observed as early as nuclear cycle 7–8), it is likely that the components involved in the early embryonic establishment of heterochromatin are deposited into the egg by the mother.

Initial evidence for a role of the rasiRNA machinery in heterochromatin formation in somatic cells came from the observation that mutations in piwi, aub, and spn-E function as weak Su
(var)s when four different variegating insertions of the white gene are examined\[117]\.
Consistent with the fact that the immediate sequence surrounding these reporters is different,
due to different genomic insertion sites and the presence of transgene arrays at two of the
reporters, the different rasiRNA components suppress variegation of the inserts to different
degrees. For example, the reporter on the arm of chromosome 4 shows a higher impact with
piwi mutations whereas a reporter in the pericentric region of the 4th chromosome is more
strongly impacted by aub and spn-E mutations. This variability of suppression is also true for
a fifth white reporter insertion, which contains a copy of the transposable element J360 in
addition to the reporter. In this case, mutations in aub show the highest level of suppression
[118]. The differences in suppression of PEV by mutations in the rasiRNA pathway become
even more complex when a white reporter inserted into a piRNA cluster at the 3R TAS is
examined. In this case, mutations in piwi cause enhancement of variegation[85]. Given these
varying effects one must keep in mind what is known about the regulation of the surrounding
heterochromatin sequence. What TEs are present in the immediate vicinity and are there piRNA
clusters at or adjacent to the insertion site? Are there other factors that might also play a role
in regulating heterochromatin at this site? One such possibility would be binding sites for the
DNA-binding protein D1, which is involved in heterochromatin formation at AT-rich satellite
DNA[119]. As more information is elucidated about heterochromatin it is apparent that it is
not homogeneous, and therefore one should hesitate to draw conclusions about
heterochromatin as a whole from the observations made with a handful of reporters.
Nonetheless, extensive investigation of these systems can certainly lead to important
discoveries.

As noted above, examination of adult female carcass tissues showed no increase in HeT-A,
copia, or I element expression in spn-E, armi, or piwi mutant lines and no change in the
chromatin structure of these TEs in spn-E mutants[100]. These observations are consistent with
a previous report of no effect on HeT-A transcript levels in spn-E, piwi, or armi mutant male
larvae. In the same study, no effect on Het-A expression with Su(var)3–9 mutations was
observed, but an increase in expression with Su(var)2–5 mutations was seen[120]. The increase
of HeT-A expression in Su(var)2–5 mutant male larvae is consistent with the effect observed
in Su(var)2–5 mutant females flies[112] and demonstrates the importance of HP1 in repressing
the expression of TEs in germline and somatic cells. The lack of any effect on HeT-A expression
in somatic cells in a Su(var)3–9 mutant may simply be due to the redundancy of HMTs as
discussed above. The silencing of a copia LTR-LacZ reporter in ovarian tissue has been
reported to be dependent upon spn-E, armi, and piwi, but silencing of this same transgene was
not affected in somatic tissues of the larvae (larval brain, haltere imaginal disc, and salivary
gland)[100]. A similar situation is observed with a GFP reporter whose expression is regulated
by a fragment of the Zam or Idefix TE[94]. These reporters are silenced in the ovaries in a
manner dependent upon piwi and the flam/COM locus. However, the silencing of these GFP
transgenes outside of the germline, in larvae, pupal and adult stages, is independent of piwi.
Interestingly, this piwi independent silencing outside the germline is still dependent upon flam/
COM, consistent with rasiRNAs playing a role in the early embryo[94].

There are several issues to consider when analyzing these seemingly contradictory results. The
regulation of a reporter inserted into a euchromatic region whose expression is controlled by
some component of a TE, such as the copia LTR-LacZ reporter or the Zam/Idefix GFP reporter,
may be different from that seen using the more traditional PEV assays. In the latter case, one
is looking at a normally euchromatic gene inserted into a heterochromatic region, which may
show different sensitivities to disruption of components in the rasiRNA system. Single copies
of P element based reporters are not silenced when inserted into the euchromatic arms, even
if the construct includes an additional TE fragment; either proximity to heterochromatic masses
or a high local TE repeat density appears to be required[118]. In addition, the requirement for
“establishment” of heterochromatin domains at the early embryo stage in the somatic tissue
may differ from regulation of TEs in the germline tissues. As most null mutations of the rasiRNA pathway members result in female sterility, parents heterozygous for the rasiRNA component are generally crossed to generate the homozygous individuals that are used in many of these assays. Therefore, a potential “half dose” of any hypothetical rasiRNA targeting complex will generally have been loaded into the embryo of the homozygous mutant individual. These complexes may be sufficient to initiate the silencing of reporters under the regulation of TE sequences, but the low dose may result in defects in heterochromatin stability, resulting in less silencing of “euchromatic gene” based reporters embedded into heterochromatin. It is also possible the there are different requirements for assembly of heterochromatin in the somatic cells and germline tissues of an organism. Not all of the effects of mutations in the rasiRNA pathway on heterochromatin formation can be explained by loss of components loaded into the embryo from the female germline, as a mutation passed through the male germline to heterozygous offspring can exhibit defects in PEV assays as well[117]. This observation argues that either there is a role for the rasiRNA components after the “establishment” of heterochromatin in the early embryo (possibly at the relaxation phase), that the onset of expression of these factors is such that they are expressed in the developing zygote before or during the stage of heterochromatin formation that they are required for, or that there is some sort of “mark” transmitted through the mutant male that causes a disruption in forming heterochromatin.

Support for a rasiRNA-based targeting complex in the early embryo comes from the observation that Piwi, the only Argonaute protein present in the nucleus, can interact with HP1α[121]. Evidence for this interaction was first obtained with a yeast-two-hybrid (Y2H) screen of an ovary cDNA library using Piwi as bait. This interaction involves contact between a PxVxL motif at the N-terminus of Piwi and the chromo shadow domain of HP1α. Piwi and HP1α can be co-immunoprecipitated from embryo nuclear extract, and co-localize in the nuclei of early embryos at the time of heterochromatin formation[121]. The functional relevance of this interaction for heterochromatin formation was demonstrated by a PEV assay where a V30A mutation in piwi (shown to impact the interaction with HP1 in the Y2H assay) resulted in less silencing at a variegating white reporter previously shown to be sensitive to piwi mutations[121]. However, there have been no reports to date of the isolation of an HP1α-Piwi complex as such, or investigations of any HP1α-coprecipitating small RNAs.

**A role for siRNA pathway components in Drosophila chromatin structure?**

Much of the data suggesting that small RNAs may be involved in heterochromatin formation in flies comes from the evidence that the rasiRNA pathway components contribute to silencing of TEs and of reporters that are located in heterochromatic regions of the fly genome. However, there are several observations that indicate cross-talk between the rasiRNA and the siRNA pathways which suggest that components of the siRNA pathway could also play a role in heterochromatin formation in flies. Factors involved in the siRNA pathway have been isolated from S2 cells in complexes capable of degrading a transcript in vitro via an RNAi based mechanism[122,123]. Complex components include siRNAs, the Argonaute family member AGO2, the Drosophila homolog of the Fragile X mental retardation protein (dFMR1), VIG (vasa intronic gene), the Drosophila homolog of p68 RNA helicase (also known as Dmp68/Rm62/Lip), TudorSN, and two ribosomal proteins, L5 and L11, along with 5S RNA[123,124]. This pathway has been shown to play a role in viral defense, as mutations in components of this pathway, as well as some components of the rasiRNA pathway (piwi, aub, armi), cause an increased sensitivity relative to wild-type flies upon viral exposure (as measured by percent survival several days post infection)[125–128]. Interestingly, components of the rasiRNA pathway, including armi, aub, and spn-E, have also been implicated in a role in classic post-transcriptional “RNAi” silencing mediated primarily via components of the siRNA pathway [78,79,81,129], suggesting some “cross-talk” between these pathways (Figure 1). Furthermore,
the recent identification of endogenous siRNAs present in both somatic and germline tissue which interact with Ago2, require Dcr2 for processing, and have homology to transposable elements supports potential “cross-talk” between these pathways in heterochromatin formation and maintenance [15–18].

Three components of the siRISC complex, Dmp68, AGO2, and dFMR1 have been linked to the formation of heterochromatin. The Drosophila p68 homologue (Lip) was first identified genetically as a suppressor of position effect variegation and of several white mutations caused by retrotransposon insertions[130]. Lip mutations also showed changes in the level of copia transcript present in the fly and alterations in the level of transcript at the white allele used in the initial screen[130]. These changes in transcript levels can potentially be explained by a role for Dmp68 in transcriptional deactivation[131]. The Dmp68 protein is an RNA helicase required for transcript clearance at sites of active transcription; mutations in Dmp68 result in RNA export defects. These functions seem to be important for deactivation of highly transcribed genes, such as the hsp70 gene following heat shock[131]. AGO2 and dFMR1 also play a role in chromosomal function during early embryogenesis[132,133]. Mutations in either AGO2 or dFMR1 cause asynchronous nuclear division cycles in the early embryo. These embryos show a mislocalization of HP1 relative to wild-type embryos and have defects in assembling centromeric heterochromatin, demonstrated by the fact that mutations in either factor show a suppression of PEV for white reporters inserted into pericentric heterochromatin. In addition, AGO2 mutants exhibit mislocalization of the centromere-specific H3 variant CID. Furthermore, both factors are required for normal development of the pole cells in the early embryo. However, none of these defects are fully penetrant, as AGO2 homozygous mutants produce viable, fertile adults, and dFMR1 null mutants are viable, although viability is reduced and there are associated neuronal defects. Possible explanations for this lack of penetrance are that the mutants used for these studies are not actually complete nulls or that there are multiple redundant pathways involved[132,133]. A model involving components of both the siRNA pathway and the rasiRNA pathway in the establishment of heterochromatin in the early embryo may be needed (Figure 3), although at present there is no direct evidence of complexes between siRNA components and heterochromatin components. Given the evidence that components of the rasiRNA pathway are also involved in viral defense in the fly and that components of the rasiRNA pathway are involved in siRNA silencing via a double-stranded RNA hairpin knock-down, it seems likely that cross-talk between these two pathways can occur when the organism is presented with different types of RNA, either viral or that expressed from TEs, that may challenge its survival (Figure 1). The detailed mechanisms behind this cross-talk remain to be worked out.

Further evidence for a link between the siRNA pathway and heterochromatin stability comes from studies showing that mutations in dcr-2, the primary Dicer of the siRNA pathway, show disorganized and multiple nucleoli, and extrachromosomal circles of ribosomal DNA (rDNA) and satellite DNA in salivary gland and diploid nuclei[134]. Interestingly, a second siRNA component, ago2, and the rasiRNA pathway components piwi, aub, and spn-E also show an increased number of nucleoli per nucleus. Similar defects are observed in lines with mutations in the heterochromatin components Su(var)3–9 or Su(var)2–5. These changes in nuclear organization correspond to a decrease in H3K9me at these loci in a dcr-2 mutant and appear to result from the generation of extra-chromosomal circular DNA. While all of the loci examined reside in heterochromatin, it is not entirely clear whether the formation of heterochromatin at the rDNA or at satellite sequences is similar to or somewhat different from that at transposable elements. However, the results do indicate a role for small RNA pathways in forming and maintaining heterochromatin at the rDNA and satellite sequences. A similar requirement for the RNAi system in maintaining nucleolar integrity has also been found in Arabidopsis[reviewed in 135]. Thus, the RNAi system is required to maintain the integrity of the genome in somatic cells in domains with a high density of repetitious sequences, most
likely through the establishment and maintenance of heterochromatin to minimize internal recombination events.

**Observations supporting a role for small RNAs in heterochromatin formation in vertebrates**

Do small RNAs play a role in directing heterochromatin formation in mammals? Several groups have identified piRNA clusters in the male germline of mice and rats that are important for spermatogenesis[23,136–140]. These clusters appear to be developmentally regulated, as some are expressed in the pachytene stage and some in pre-pachytene stages. The clusters expressed in pre-pachytene are more transposon-rich and have been implicated in transposon control in the germline, presumably using a mechanism similar to the rasiRNA pathway in flies [23]. Additional support for a role of the RNAi machinery in transposon control in mammals comes from disruption of the genes for the mouse piwi homologues Mili or Miwi2, which results in derepression and loss of methylation of transposons in the male germline[23,25]. Analysis of mouse oocytes found small RNA of the siRNA class (~20–24nt) derived from retroelements [138], implicating this second class of small RNAs in transposon control (at least in the oocyte) as well. Initial studies on mouse embryonic stem cells with a deletion of the Dicer gene reported an increased accumulation of centromeric transcripts from the major and minor satellite repeats; however no changes in DNA methylation or H3K9 methylation were observed at these loci[141]. More recent studies found changes in DNA methylation in Dicer mutants, but these effects appear to be mediated through the microRNA pathway via regulation of DNA methyltransferases by Rbl2 and the mir-290 cluster, indicating an indirect role for these small RNAs in epigenetic regulation[142,143]. Thus, there is accumulating evidence of a role for small RNAs in controlling repetitious sequences in the germline, but as yet little or no evidence for a direct role in somatic cells.

**A model for heterochromatin formation in flies**

Given the overall body of evidence it seems highly likely that small RNAs are playing a role in directing heterochromatin formation in Drosophila, but clearly there are many details that remain to be elucidated, and additional evidence is required to confirm this role. Two pathways could be operating in parallel in flies, one in the germline and a second in the early embryo, the latter required to set up heterochromatin structure in somatic nuclei and maintain it throughout development of the organism. However, it seems likely that components of a germline system are present, and could help direct the embryonic/somatic pathway, potentially by depositing small RNAs and associated proteins into the embryo. It also appears likely that a transcription-based mechanism could be used, similar to that in *S. pombe*, but potentially without the cis restriction (Figure 3). Exactly what “RNAi” machinery is used may depend upon the type of transposable element/repeat to be silenced and what tissue/cell type (germline or somatic) is examined, as these factors may influence which components are recruited. In the germline, it appears that the rasiRNA pathway components play the primary, if not only, role in regulating TE expression and mobilization, likely through a post-transcriptional mechanism but also by controlling the formation/maintenance of heterochromatin, whereas in the early embryo both the rasiRNA pathway and siRNA pathway components may play a role in regulating TEs and directing heterochromatin formation. As preserving germline integrity is crucial for the success of future generations, TE expression/mobilization in this tissue may be under more stringent control than in somatic tissue. However, regulation of TEs in somatic tissues is presumably required to prevent mutations caused by mobilization as well as destabilization of heterochromatin that could decrease survival and hence the ability of the organism to propagate. As formation of heterochromatin is critical for basic cellular processes in the organism, it is quite possible that there are additional pathways outside of a small RNA based mechanism, such as the DNA-protein interactions observed at the MAT loci in *S.*
pombe, which direct heterochromatin formation. It is likely that some of the primary players, such as HP1α and H3K9me, will function as the “end-point” for multiple pathways in setting up heterochromatin in the organism. Once heterochromatin is formed, it appears to be sufficient to maintain TE in a silent state, as mutations in the rasiRNA pathway no longer have a major impact. The requirement for the rasiRNA pathway in the germline may indicate that heterochromatin formation is less efficacious there; the dynamic state of the germline could be designed to inoculate the embryo against viral challenges later in life. The siRNA/rasiRNA system is complex, as required to fill its various roles: defense against invading viruses, protection against endogenous transposable elements, maintenance of genome stability in regions with high levels of repetitious sequences, accomplished in both germ line and somatic cells. Sorting out the various pathways involved, including cross-talk between components will take many more years of effort by numerous investigators.

Acknowledgements

We thank Brent Brower-Toland, Nicole Riddle, and Elena Gracheva for critical comments on this manuscript. K. Huisinga is supported by a Kirschstein NRSA Postdoctoral Fellowship from NIH; work from the Elgin lab cited above is supported by NIH grants R01 GM68388 and GM073190 to CRE.

References


Figure 1.
Overview of small RNA pathways in Drosophila. Members of the RNAse III family of nucleases are in light red, members of the DRB family are in light blue, and members of the Argonaute family of proteins are in green/yellow/orange. The cross-talk between Dcr-1 and Dcr-2\cite{4} and potentially between the siRNA and rasiRNA pathways, discussed in the text, is indicated.
Figure 2.
Schematic model for cell-cycle dependent regulation of heterochromatin formation at *S. pombe* pericentric regions. The self-reinforcing loop of heterochromatin formation is illustrated during S-phase, when transcription of the *dg-dh* region occurs. Adapted from figures in [55, 63, 64]. The diagram at M phase has been simplified to emphasize the transition in Swi6 binding.
Figure 3.
Model for small RNA mediated regulation of TE silencing in the Drosophila germline and potential links to heterochromatin formation in germline and early embryo. The proposed ping-pong mechanism of TE regulation in the germline is illustrated on the left. Links between rasiRNA machinery (much of which is present in the nuage of the female germline) and heterochromatin formation are diagramed. Other RNAi components implicated in heterochromatin formation or maintenance are listed in Table 1.
### Table 1

**Regulation of TE expression in Ovaries**

Expression of various transposable elements as determined by Northern, RT-PCR, or *in situ* hybridization in the ovaries of flies carrying mutations in various components of small RNA processing systems relative to that observed in the ovaries of wild-type flies. An asterisk indicates analysis in larval gonads. Discrepancies between different publications are highlighted; N/A = not analyzed.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Cellular localization</th>
<th>Tissue type</th>
<th>Mutant viability/sterility</th>
<th>Non-LTR retrotransposon</th>
<th>LTR retrotransposon</th>
<th>Repeat locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>hobo</td>
<td>Argonaute</td>
<td>Nucleus</td>
<td>Somatic and germline</td>
<td>Female sterile, male sterile</td>
<td>Yes, small&lt;sup&gt;1,2,6&lt;/sup&gt; No&lt;sup&gt;1,6&lt;/sup&gt;</td>
<td>HeT-A Yes, large&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Mst40</td>
</tr>
<tr>
<td>sab</td>
<td>Argonaute</td>
<td>Nuage-cyto</td>
<td>Germline</td>
<td>Female sterile</td>
<td>Yes, large&lt;sup&gt;1,10,12&lt;/sup&gt; Yes&lt;sup&gt;5,8,10,12&lt;/sup&gt;,&lt;sup&gt;13&lt;/sup&gt; no (2, heterozygotes)</td>
<td>TART Yes-mod&lt;sup&gt;14&lt;/sup&gt;</td>
<td>F-element No&lt;sup&gt;1, yes-mod&lt;sup&gt;14&lt;/sup&gt;&lt;/sup&gt;</td>
</tr>
<tr>
<td>AGO3</td>
<td>Argonaute</td>
<td>Nuage-cyto</td>
<td>Germline</td>
<td></td>
<td>Yes, medium&lt;sup&gt;6,10,12&lt;/sup&gt; Yes&lt;sup&gt;5,8,10,12&lt;/sup&gt;,&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Y. slight&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>spn-E</td>
<td>RNA helicase</td>
<td>N/A</td>
<td>Germline (female)</td>
<td>Female sterile</td>
<td>Yes, medium&lt;sup&gt;6,10,12&lt;/sup&gt; Yes&lt;sup&gt;5,8,10,12&lt;/sup&gt;,&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Y. slight&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>wnt</td>
<td>RNA helicase</td>
<td>Nuage-cyto</td>
<td>Germline</td>
<td>Female sterile</td>
<td>Yes, medium&lt;sup&gt;6,10,12&lt;/sup&gt; Yes&lt;sup&gt;5,8,10,12&lt;/sup&gt;,&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Y. slight&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>AG02</td>
<td>Argonaute</td>
<td>Nuage-cyto</td>
<td>Somatic</td>
<td>Viable</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td>V. slight&lt;sup&gt;4&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Dcr-2</td>
<td>Dicer</td>
<td>Cyto</td>
<td>Somatic</td>
<td>Viable</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>R2D2</td>
<td>DRB</td>
<td>Cyto</td>
<td>Somatic</td>
<td>Viable</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>spn-E()</td>
<td>Dicer</td>
<td>Cyto</td>
<td>Somatic</td>
<td>Viable</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Loqs</td>
<td>DRB</td>
<td>Cyto</td>
<td>Somatic</td>
<td>Viable</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td>No&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>zuc</td>
<td>Nuclease</td>
<td>Nuage-cyto</td>
<td>Germline</td>
<td>Female sterile</td>
<td>Yes&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>spn-E</td>
<td>Nuclease</td>
<td>Nuage-cyto</td>
<td>Germline</td>
<td>Female sterile</td>
<td>Yes&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cuff</td>
<td>Homologue of yeast Gene Rail</td>
<td>Partial nuage and cyto</td>
<td>Germline</td>
<td>Female sterile</td>
<td>Yes&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>musl</td>
<td>Partial potential HMG-box</td>
<td>Shuttle-nuage and cyto and germline</td>
<td>Somatic and germline</td>
<td>Female sterile</td>
<td>Yes&lt;sup&gt;8&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>hrimp</td>
<td>Predicted Tudor domain</td>
<td>Nuage</td>
<td>Germline</td>
<td>Female sterile</td>
<td>Yes&lt;sup&gt;8&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;8&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>


